Non-intrusive Termination of Noisy Optimization

Jeffrey Larson
University of Colorado Denver

Stefan Wild
Argonne National Laboratory

FEBRUARY 24, 2012

Problem Setting

When should you terminate algorithms solving

$$\min_x f(x) : \mathbb{R}^n \to \mathbb{R}$$

when f is

- Computationally expensive
- There is noise in the computation of f

Problem Setting

When should you terminate algorithms solving

$$\min_{x} f(x) : \mathbb{R}^n \to \mathbb{R}$$

when f is

- Computationally expensive
- There is noise in the computation of f

Practitioners typically stop the optimization when:

- A measure of criticality is satisfied (e.g., gradient norm, mesh size)
- The computational budget is satisfied (e.g., number of evaluations, wall clock time)

Problem Setting

When should you terminate algorithms solving

$$\min_{x} f(x) : \mathbb{R}^n \to \mathbb{R}$$

when f is

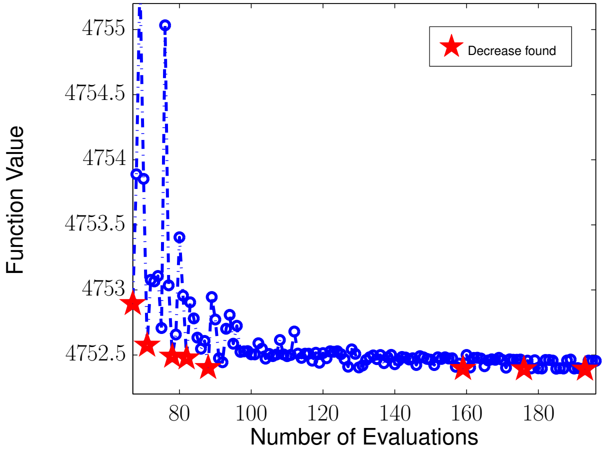
- Computationally expensive
- There is noise in the computation of f

Definition

This is an attempt to solve the true problem:

$$\begin{array}{ll} \min\limits_{t} & \text{Computational expense}(t) \\ \text{s.t.} & \text{Acceptable accuracy of the solution}(t), \end{array}$$

Example from Nuclear Physics



Quotes

. . . no set of termination criteria is suitable for all optimization problems and all methods.

- P. Gill, W. Murray, M. Wright (Practical Optimization)

. . . it is believed that it is impossible to choose such a convergence criterion which is effective for the most general function . . . so a compromise has to be made between stopping the iterative procedure too soon and calculating f an unnecessarily large number of times.

M. Powell (1964)

Quotes

. . . no set of termination criteria is suitable for all optimization problems and all methods.

- P. Gill, W. Murray, M. Wright (Practical Optimization)

. . . it is believed that it is impossible to choose such a convergence criterion which is effective for the most general function . . . so a compromise has to be made between stopping the iterative procedure too soon and calculating f an unnecessarily large number of times.

M. Powell (1964)

Modifications for Noisy Function

1. Functions with stochastic noise, $\mathrm{Var}\{f(x)\} > 0$,
   - UOBYQA, DIRECT, and Nelder-Mead methods have all been modified in the literature to repeatedly sample points.
   - Some adjust the maximum number of replications based on the noise level.
   - Termination was still based on traditional measures:
     - points clustered together
     - no decrease in the best function value

2. Functions with deterministic noise, (iterative methods, round-off error)
   - Kelley- proposes a restart technique for Nelder-Mead when low-level noise is present, but terminates independent of the noise.
   - Gramacy et al. - stops a treed Gaussian process when the maximum improvement statistic is sufficiently small.
   - Neumaier et al. - Suggest stopping SNOBFIT when the best point has not changed for a number of consecutive iterates.

Modifications for Noisy Function

1. Functions with stochastic noise, $\mathrm{Var}\{f(x)\} > 0$,
   - UOBYQA, DIRECT, and Nelder-Mead methods have all been modified in the literature to repeatedly sample points.
   - Some adjust the maximum number of replications based on the noise level.
   - Termination was still based on traditional measures:
     - points clustered together
     - no decrease in the best function value
2. Functions with deterministic noise, (iterative methods, round-off error)
   - Kelley- proposes a restart technique for Nelder-Mead when low-level noise is present, but terminates independent of the noise.
   - Gramacy et al. - stops a treed Gaussian process when the maximum improvement statistic is sufficiently small.
   - Neumaier et al. - Suggest stopping SNOBFIT when the best point has not changed for a number of consecutive iterates.

Modifications for Noisy Function

1. Functions with stochastic noise, $\mathrm{Var}\left\{f(x)\right\} > 0$,
   - UOBYQA, DIRECT, and Nelder-Mead methods have all been modified in the literature to repeatedly sample points.
   - Some adjust the maximum number of replications based on the noise level.
   - Termination was still based on traditional measures:
     - points clustered together
     - no decrease in the best function value

2. Functions with deterministic noise, (iterative methods, round-off error)
   - Kelley- proposes a restart technique for Nelder-Mead when low-level noise is present, but terminates independent of the noise.
   - Gramacy et al. - stops a treed Gaussian process when the maximum improvement statistic is sufficiently small.
   - Neumaier et al. - Suggest stopping SNOBFIT when the best point has not changed for a number of consecutive iterates.

Desirable Test Properties

For a sequence of points and function values

$$\{x_1, \cdots, x_m\} \subseteq \mathbb{R}^n, \{f_1, \cdots, f_m\} \in \mathbb{R}, \mathcal{F}_i = \{(x_1, f_1), \cdots, (x_i, f_i)\}$$

produced by a local minimization solver, it is preferable if the termination test is:

- Algorithm independent
  - Uses only the x_i and f_i.
- Shift and scale invariant in f
  - Stops sequences $\{f_i\}$ and $\{\alpha f_i + \beta\}$ at the same point.

Useful notation: Let $f_i^* = \min_{1 \le j \le i} \{f_j\}$ and x_i^* be the corresponding point.

Desirable Test Properties

For a sequence of points and function values

$$\{x_1, \cdots, x_m\} \subseteq \mathbb{R}^n, \{f_1, \cdots, f_m\} \in \mathbb{R}, \mathcal{F}_i = \{(x_1, f_1), \cdots, (x_i, f_i)\}$$

produced by a local minimization solver, it is preferable if the termination test is:

- Algorithm independent
  - Uses only the x_i and f_i.
- Shift and scale invariant in f
  - Stops sequences $\{f_i\}$ and $\{\alpha f_i + \beta\}$ at the same point.

Useful notation: Let $f_i^* = \min\limits_{1 \leq j \leq i} \{f_j\}$ and x_i^* be the corresponding point.

Estimate of the Noise Level

Let $\hat{\varepsilon}_{i_r}$ be the relative noise in f_i.

For stochastic noise, $\hat{\varepsilon}_{i_r} = \dfrac{\sqrt{\mathrm{Var}\{f(x_i)\}}}{E\{|f(x_i)|\}}$, in which case, for $\alpha > 0$:

$$\hat{\varepsilon}_{i_r} = \frac{\alpha\sqrt{\mathrm{Var}\left\{f(x_i)\right\}}}{\alpha E\left\{|f(x_i)|\right\}} = \frac{\sqrt{\alpha^2\mathrm{Var}\left\{f(x_i)\right\}}}{\alpha E\left\{|f(x_i)|\right\}} = \frac{\sqrt{\mathrm{Var}\left\{\alpha f(x_i)\right\}}}{E\left\{|\alpha f(x_i)|\right\}}.$$

$\hat{\varepsilon}_{i_r}$ is scale invariant.

$$\hat{\varepsilon}_{i_r} E\left\{|f(x_i)|\right\} = \sqrt{\mathrm{Var}\left\{f(x_i)\right\}} = \sqrt{\mathrm{Var}\left\{f(x_i) + \beta\right\}}$$

$\hat{\varepsilon}_{i_r} E\left\{|f(x_i)|\right\}$ is shift invariant.

For deterministic noise, invariance depends on the methods used to obtain $\hat{\varepsilon}_{i_r}$ and $\hat{\varepsilon}_{i_r}|f_i|$. For one such method, see Moré & Wild (2011).

Estimate of the Noise Level

Let $\hat{\varepsilon}_{i_r}$ be the relative noise in f_i.

For stochastic noise, $\hat{\varepsilon}_{i_r} = \frac{\sqrt{\mathrm{Var}\{f(x_i)\}}}{E\{|f(x_i)|\}}$, in which case, for $\alpha > 0$:

$$\hat{\varepsilon}_{i_r} = \frac{\alpha\sqrt{\mathrm{Var}\left\{f(x_i)\right\}}}{\alpha E\left\{|f(x_i)|\right\}} = \frac{\sqrt{\alpha^2\mathrm{Var}\left\{f(x_i)\right\}}}{\alpha E\left\{|f(x_i)|\right\}} = \frac{\sqrt{\mathrm{Var}\left\{\alpha f(x_i)\right\}}}{E\left\{|\alpha f(x_i)|\right\}}.$$

$\hat{\varepsilon}_{i_r}$ is scale invariant.

$$\hat{\varepsilon}_{i_r} E\left\{|f(x_i)|\right\} = \sqrt{\mathrm{Var}\left\{f(x_i)\right\}} = \sqrt{\mathrm{Var}\left\{f(x_i) + \beta\right\}}$$

$\hat{\varepsilon}_{i_r} E\left\{|f(x_i)|\right\}$ is shift invariant.

For deterministic noise, invariance depends on the methods used to obtain $\hat{\varepsilon}_{i_r}$ and $\hat{\varepsilon}_{i_r}|f_i|$. For one such method, see Moré & Wild (2011).

Estimate of the Noise Level

Let $\hat{\varepsilon}_{i_r}$ be the relative noise in f_i.

For stochastic noise, $\hat{\varepsilon}_{i_r} = \frac{\sqrt{\mathrm{Var}\{f(x_i)\}}}{E\{|f(x_i)|\}}$, in which case, for $\alpha > 0$:

$$\hat{\varepsilon}_{i_r} = \frac{\alpha\sqrt{\mathrm{Var}\left\{f(x_i)\right\}}}{\alpha E\left\{|f(x_i)|\right\}} = \frac{\sqrt{\alpha^2\mathrm{Var}\left\{f(x_i)\right\}}}{\alpha E\left\{|f(x_i)|\right\}} = \frac{\sqrt{\mathrm{Var}\left\{\alpha f(x_i)\right\}}}{E\left\{|\alpha f(x_i)|\right\}}.$$

$\hat{\varepsilon}_{i_r}$ is scale invariant.

$$\hat{\varepsilon}_{i_r} E\left\{|f(x_i)|\right\} = \sqrt{\mathrm{Var}\left\{f(x_i)\right\}} = \sqrt{\mathrm{Var}\left\{f(x_i) + \beta\right\}}$$

$\hat{\varepsilon}_{i_r} E\left\{|f(x_i)|\right\}$ is shift invariant.

For deterministic noise, invariance depends on the methods used to obtain $\hat{\varepsilon}_{i_r}$ and $\hat{\varepsilon}_{i_r}|f_i|$. For one such method, see Moré & Wild (2011).

Estimate of the Noise Level

Let $\hat{\varepsilon}_{i_r}$ be the relative noise in f_i.

For stochastic noise, $\hat{\varepsilon}_{i_r} = \frac{\sqrt{\mathrm{Var}\{f(x_i)\}}}{E\{|f(x_i)|\}}$, in which case, for $\alpha > 0$:

$$\hat{\varepsilon}_{i_r} = \frac{\alpha\sqrt{\mathrm{Var}\left\{f(x_i)\right\}}}{\alpha E\left\{|f(x_i)|\right\}} = \frac{\sqrt{\alpha^2\mathrm{Var}\left\{f(x_i)\right\}}}{\alpha E\left\{|f(x_i)|\right\}} = \frac{\sqrt{\mathrm{Var}\left\{\alpha f(x_i)\right\}}}{E\left\{|\alpha f(x_i)|\right\}}.$$

$\hat{\varepsilon}_{i_r}$ is scale invariant.

$$\hat{\varepsilon}_{i_r} E\left\{|f(x_i)|\right\} = \sqrt{\mathrm{Var}\left\{f(x_i)\right\}} = \sqrt{\mathrm{Var}\left\{f(x_i) + \beta\right\}}$$

$\hat{\varepsilon}_{i_r} E\left\{|f(x_i)|\right\}$ is shift invariant.

For deterministic noise, invariance depends on the methods used to obtain $\hat{\varepsilon}_{i_r}$ and $\hat{\varepsilon}_{i_r}|f_i|$. For one such method, see Moré & Wild (2011).

Tests on Function Values, ϕ_1

For $\nu_{\mathcal{F}_i}, \mu \in \mathbb{R}_+,\ \kappa \in \mathbb{N}$. Let $\hat{\varepsilon}_{i_r}$ be an estimate for the relative noise level of f_i.

$$\phi_1\left(\mathcal{F}_i; \nu_{\mathcal{F}_i}, \kappa, \mu\right) \text{ stops when } \frac{f^*_{i-\kappa+1} - f^*_i}{\kappa} \le \mu \left|f^*_i\right| \nu_{\mathcal{F}_i}$$

- If $\nu_{\mathcal{F}_i} = 1$: stop when the average relative change in the best function value over the last κ evaluations is less than μ. (scale invariant)
- If $\nu_{\mathcal{F}_i} = \hat{\varepsilon}_{i_r}$: stop when the average relative change in f^* is over the last κ evaluations is less than a factor of μ times the relative noise. (shift and scale invariant)

Tests on Function Values, ϕ_1

For $\nu_{\mathcal{F}_i}, \mu \in \mathbb{R}_+,\ \kappa \in \mathbb{N}$. Let $\hat{\varepsilon}_{i_r}$ be an estimate for the relative noise level of f_i.

$$\phi_1\left(\mathcal{F}_i; \nu_{\mathcal{F}_i}, \kappa, \mu\right) \text{ stops when } \frac{f^*_{i-\kappa+1} - f^*_i}{\kappa} \leq \mu \left|f^*_i\right| \nu_{\mathcal{F}_i}$$

- If $\nu_{\mathcal{F}_i} = 1$: stop when the average relative change in the best function value over the last κ evaluations is less than μ. (scale invariant)
- If $\nu_{\mathcal{F}_i} = \hat{\varepsilon}_{i_r}$: stop when the average relative change in f^* is over the last κ evaluations is less than a factor of μ times the relative noise. (shift and scale invariant)

Tests on Function Values, ϕ_2

For $\nu_{\mathcal{F}_i}, \mu \in \mathbb{R}_+,\ \kappa \in \mathbb{N}$. Let $\hat{\varepsilon}_{i_r}$ be an estimate for the relative noise level of f_i.

$$\phi_2\left(\mathcal{F}_i; \nu_{\mathcal{F}_i}, \kappa, \mu\right) \text{ stops when } \max_{i-\kappa+1 \le j \le i} \left|f_j - f_i^*\right| \le \mu \left|f_i^*\right| \nu_{\mathcal{F}_i}$$

- If $\nu_{\mathcal{F}_i} = 1$: stop when κ consecutive function values are within μ of $\left|f_i^*\right|$. (scale invariant)
- If $\nu_{\mathcal{F}_i} = \hat{\varepsilon}_{i_r}$: stop when the maximum absolute change in f over the last κ evaluations is less than a factor of μ times the noise level at f_i^*. (shift and scale invariant)

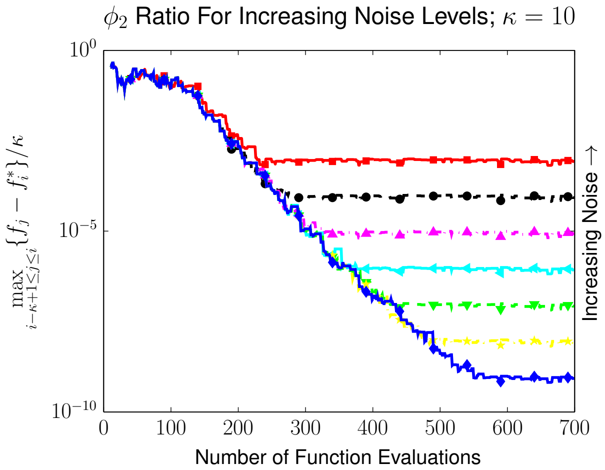
Tests on Function Values, ϕ_2

For $\nu_{\mathcal{F}_i}, \mu \in \mathbb{R}_+,\ \kappa \in \mathbb{N}$. Let $\hat{\varepsilon}_{i_r}$ be an estimate for the relative noise level of f_i.

$$\phi_2\left(\mathcal{F}_i; \nu_{\mathcal{F}_i}, \kappa, \mu\right) \text{ stops when } \max_{i-\kappa+1 \le j \le i} \left|f_j - f_i^*\right| \le \mu \left|f_i^*\right| \nu_{\mathcal{F}_i}$$

- If $\nu_{\mathcal{F}_i} = 1$: stop when κ consecutive function values are within μ of $\left|f_i^*\right|$.
  (scale invariant)
- If $\nu_{\mathcal{F}_i} = \hat{\varepsilon}_{i_r}$: stop when the maximum absolute change in f over the last κ evaluations is less than a factor of μ times the noise level at f_i^*.
  (shift and scale invariant)

Dependence on the Noise Level



Tests on x Values

For $\mu \in \mathbb{R}_+$, $\kappa \in \mathbb{N}$.

- $\phi_3(\mathcal{F}_i; \kappa, \mu)$ stops when $\max_{i-\kappa+1 \leq j,k \leq i} \|x_j - x_k\| \leq \mu$
  - Stop when κ consecutive x-values are within a distance μ of each other.
- $\phi_4(\mathcal{F}_i; \kappa, \mu)$ stops when $\max_{i-\kappa+1 \leq j \leq i} \left\|x_j^* - x_i^*\right\| \leq \mu$
  - Stop when κ consecutive x_i^*-values are within a distance μ of each other.
- These tests are only shift (scale) invariant if the procedure which generates the $\{x_i\}$ is shift (scale) invariant.

Tests on x Values

For $\mu \in \mathbb{R}_+,\ \kappa \in \mathbb{N}$.

- $\phi_3\left(\mathcal{F}_i; \kappa, \mu\right)$ stops when $\displaystyle\max_{i-\kappa+1 \le j,k \le i} \left\|x_j - x_k\right\| \le \mu$
  - Stop when κ consecutive x-values are within a distance μ of each other.
- $\phi_4\left(\mathcal{F}_i; \kappa, \mu\right)$ stops when $\displaystyle\max_{i-\kappa+1 \le j \le i} \left\|x_j^* - x_i^*\right\| \le \mu$
  - Stop when κ consecutive x_i^*-values are within a distance μ of each other.
- These tests are only shift (scale) invariant if the procedure which generates the $\{x_i\}$ is shift (scale) invariant.

Tests on x Values

For $\mu \in \mathbb{R}_+$, $\kappa \in \mathbb{N}$.

- $\phi_3 (\mathcal{F}_i; \kappa, \mu)$ stops when $\max\limits_{i-\kappa+1 \leq j,k \leq i} \left\| x_j - x_k \right\| \leq \mu$
  - Stop when κ consecutive x-values are within a distance μ of each other.
- $\phi_4 (\mathcal{F}_i; \kappa, \mu)$ stops when $\max\limits_{i-\kappa+1 \leq j \leq i} \left\| x_j^* - x_i^* \right\| \leq \mu$
  - Stop when κ consecutive x_i^*-values are within a distance μ of each other.
- These tests are only shift (scale) invariant if the procedure which generates the $\{x_i\}$ is shift (scale) invariant.

Comparison Test

As a point of comparison, we define the test

- $\phi_5(\mathcal{F}_i; \kappa)$ to stop after κ iterations

This test is trivially shift and scale invariant.

Problem Set

53 problems of the form:

$$f(x) = 1 + (1 + \sigma g(x)) \sum_{i=1}^{m} F_i^s(x)^2,$$

- For stochastic noise

$$\mathrm{Var}\left\{g(x)\right\} = 1$$

- For deterministic noise

$$g(x) = \xi(x)(4\xi(x)^2 - 3))$$

$$\xi(x) = 0.9 \sin(100\|x\|_1) \cos(100\|x\|_\infty) + 0.1 \cos(\|x\|_2).$$

Problem Set

53 problems of the form:

$$f(x) = 1 + (1 + \sigma g(x)) \sum_{i=1}^{m} F_i^s(x)^2,$$

- For stochastic noise

$$\mathrm{Var}\left\{g(x)\right\} = 1$$

- For deterministic noise

$$g(x) = \xi(x)(4\xi(x)^2 - 3))$$

$$\xi(x) = 0.9 \sin(100\|x\|_1) \cos(100\|x\|_\infty) + 0.1 \cos(\|x\|_2).$$

Problem Set

We have 6 algorithms from the following classes:

1. Nelder-Mead implementations
2. Pattern search methods
3. Model-based methods
4. . . . and methods which cross these classes

that we ran on all 53 problems, leaving us with 318 algorithm runs to form $\mathcal{P}$.

For each termination test t and $p \in \mathcal{P}$, let

$$i^*_{p,t}$$

be the number of function values required to satisfy t on problem p.

Problem Set

We have 6 algorithms from the following classes:

1. Nelder-Mead implementations
2. Pattern search methods
3. Model-based methods
4. . . . and methods which cross these classes

that we ran on all 53 problems, leaving us with 318 algorithm runs to form $\mathcal{P}$.

For each termination test t and $p \in \mathcal{P}$, let

$$i^*_{p,t}$$

be the number of function values required to satisfy t on problem p.

Measures of Quality in a Stopping Point

- Accuracy: How far from the best point does the test stop?

$$\frac{f^*_{i^*_{p,t}} - f^*_{i_{\max}}}{f^*_{i^*_{p,t}}} \quad \text{if} \quad i^*_{p,t} < i_{\max}$$

- Performance: Could the test have stopped sooner?

  Given a collection of tests $\mathcal{T}$, what $t \in \mathcal{T}$ stops when

$$f^*_{i^*_{p,t}} - f^*_{i_{\max}} \leq \left| f^*_{i^*_{p,t}} \right| \hat{\varepsilon}_{i_r}$$

  with the smallest $i^*_{p,t}$?

Measures of Quality in a Stopping Point

- Accuracy: How far from the best point does the test stop?

$$\frac{f^*_{i^*_{p,t}} - f^*_{i_{\max}}}{f^*_{i^*_{p,t}}} \quad \text{if} \quad i^*_{p,t} < i_{\max}$$

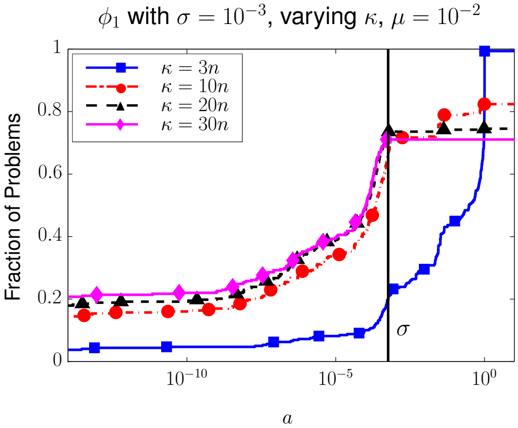
- Performance: Could the test have stopped sooner?

  Given a collection of tests $\mathcal{T}$, what $t \in \mathcal{T}$ stops when

$$f^*_{i^*_{p,t}} - f^*_{i_{\max}} \leq \left| f^*_{i^*_{p,t}} \right| \hat{\varepsilon}_{i_r}$$

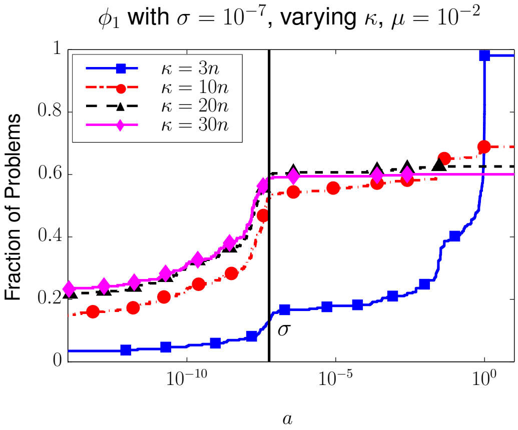
  with the smallest $i^*_{p,t}$?

Accuracy Profiles



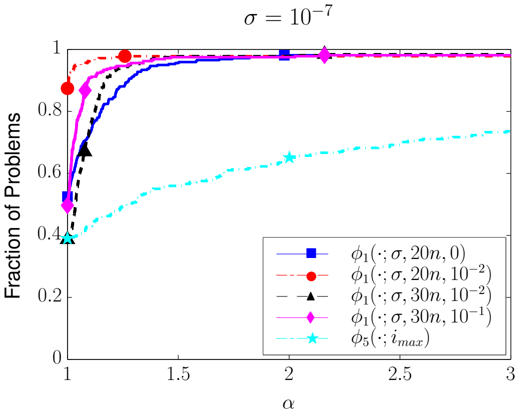
$\phi_1(\mathcal{F}_i; \sigma, \kappa, \mu)$ stops when $\frac{f^*_{i-\kappa+1} - f^*_i}{\kappa} \le \mu \left| f^*_i \right| \sigma$

Accuracy Profiles



$\phi_1\left(\mathcal{F}_i; \sigma, \kappa, \mu\right)$ stops when $\frac{f^*_{i-\kappa+1} - f^*_i}{\kappa} \leq \mu \left| f^*_i \right| \sigma$

Performance Profiles



$\phi_1\left(\mathcal{F}_i;\sigma,\kappa,\mu\right)$ stops when $\frac{f^*_{i-\kappa+1}-f^*_i}{\kappa} \leq \mu\left|f^*_i\right|\sigma$

Other test

We performed similar analysis on the other families of tests and found the best:

- $\phi_1(\cdot, \cdot, 20n, 10^{-2})$ Stop when
$$\frac{f^*_{i-\kappa+1} - f^*_i}{20n} \leq 0.01 \left| f^*_i \right| \hat{\varepsilon}_{i_r}$$
- $\phi_2(\cdot, \cdot, 10n, 10)$ Stop when
$$\max_{i-10n+1 \leq j \leq i} \left| f_j - f^*_i \right| \leq 10 \left| f^*_i \right| \hat{\varepsilon}_{i_r}$$
- $\phi_3(\cdot, n, 10^{-7})$ Stop when
$$\max_{i-n+1 \leq j,k \leq i} \left\| x_j - x_k \right\| \leq 10^{-7}$$
- $\phi_4(\cdot, 20n, 10^{-7})$ Stop when
$$\max_{i-20n+1 \leq j \leq i} \left\| x^*_j - x^*_i \right\| \leq 10^{-7}$$

Other test

We performed similar analysis on the other families of tests and found the best:

- $\phi_1(\cdot,\cdot,20n,10^{-2})$ Stop when
$$\frac{f^*_{i-\kappa+1}-f^*_i}{20n} \leq 0.01\,|f^*_i|\,\hat{\varepsilon}_{i_r}$$
- $\phi_2(\cdot,\cdot,10n,10)$ Stop when
$$\max_{i-10n+1\leq j\leq i} |f_j - f^*_i| \leq 10\,|f^*_i|\,\hat{\varepsilon}_{i_r}$$
- $\phi_3(\cdot,n,10^{-7})$ Stop when
$$\max_{i-n+1\leq j,k\leq i} \|x_j - x_k\| \leq 10^{-7}$$
- $\phi_4(\cdot,20n,10^{-7})$ Stop when
$$\max_{i-20n+1\leq j\leq i} \|x^*_j - x^*_i\| \leq 10^{-7}$$

Other test

We performed similar analysis on the other families of tests and found the best:

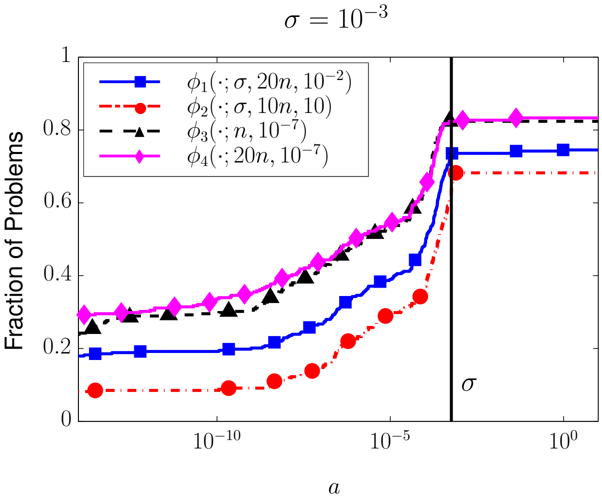
- $\phi_1(\cdot, \cdot, 20n, 10^{-2})$ Stop when
$$\frac{f^*_{i-\kappa+1} - f^*_i}{20n} \leq 0.01 \left| f^*_i \right| \hat{\varepsilon}_{i_r}$$
- $\phi_2(\cdot, \cdot, 10n, 10)$ Stop when
$$\max_{i-10n+1 \leq j \leq i} \left| f_j - f^*_i \right| \leq 10 \left| f^*_i \right| \hat{\varepsilon}_{i_r}$$
- $\phi_3(\cdot, n, 10^{-7})$ Stop when
$$\max_{i-n+1 \leq j,k \leq i} \left\| x_j - x_k \right\| \leq 10^{-7}$$
- $\phi_4(\cdot, 20n, 10^{-7})$ Stop when
$$\max_{i-20n+1 \leq j \leq i} \left\| x^*_j - x^*_i \right\| \leq 10^{-7}$$

Other test

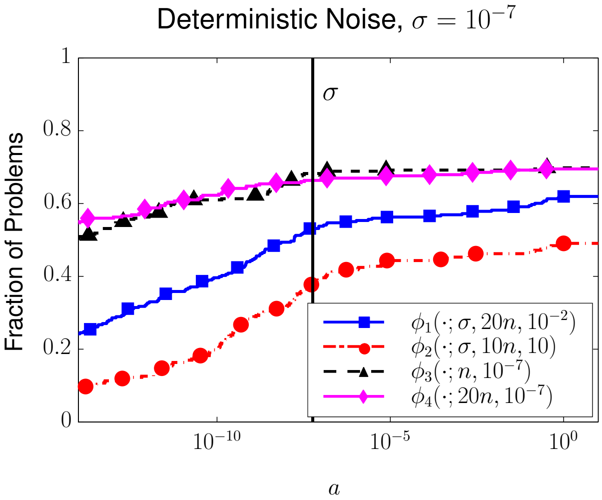
We performed similar analysis on the other families of tests and found the best:

- $\phi_1(\cdot,\cdot,20n,10^{-2})$ Stop when
$$\frac{f^*_{i-\kappa+1}-f^*_i}{20n} \leq 0.01\left|f^*_i\right|\hat{\varepsilon}_{i_r}$$
- $\phi_2(\cdot,\cdot,10n,10)$ Stop when
$$\max_{i-10n+1\leq j\leq i}\left|f_j - f^*_i\right| \leq 10\left|f^*_i\right|\hat{\varepsilon}_{i_r}$$
- $\phi_3(\cdot,n,10^{-7})$ Stop when
$$\max_{i-n+1\leq j,k\leq i}\left\|x_j - x_k\right\| \leq 10^{-7}$$
- $\phi_4(\cdot,20n,10^{-7})$ Stop when
$$\max_{i-20n+1\leq j\leq i}\left\|x^*_j - x^*_i\right\| \leq 10^{-7}$$

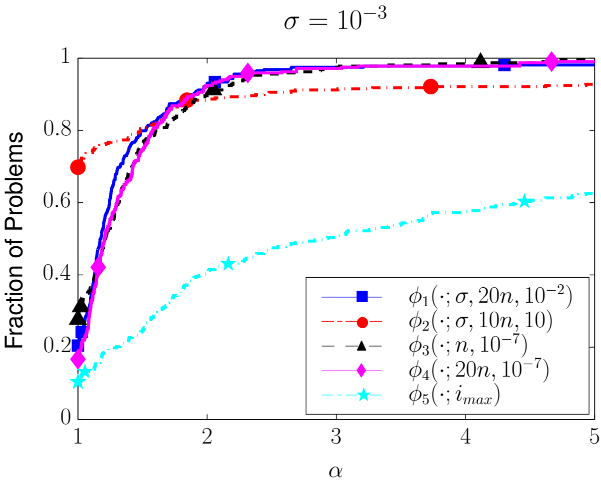
Most Accurate Tests



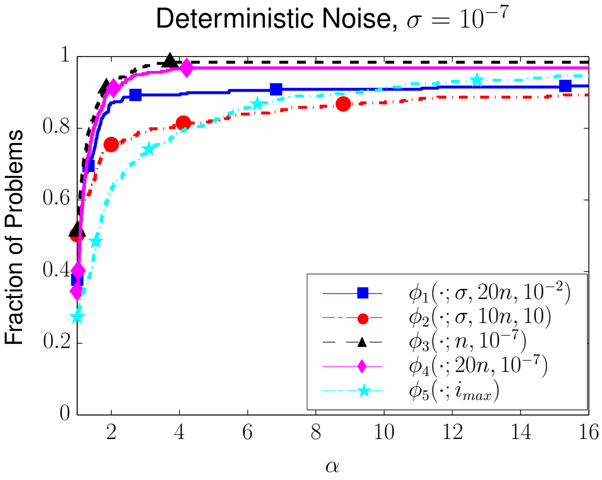
Most Accurate Tests



Performance Profiles for Most Accurate Tests



Performance Profiles for Most Accurate Tests



Recommendations for termination tests

Test	κ	μ	Interpretation of Stopping Rule
ϕ_1	$\approx 20n$	≈ 0.01	Stop when the average relative decrease in the best function value over the last $20n$ function evaluations is less than one-hundredth of the relative noise level.
ϕ_2	$\approx 10n$	≈ 10	Stop when the last $10n$ function evaluations are within 10 times the absolute noise level of the best function value.
ϕ_3	$\approx n$	$\approx 10^{-7}$	Stop when the last n points evaluated are within a distance of 10^{-7} of each other.
ϕ_4	$\approx 20n$	$\approx 10^{-7}$	Stop when the best point hasn't moved more a distance of 10^{-7} for $20n$ evaluations.

Final Comments:

- Tests using knowledge of the noise are better, especially as the noise level increases.

- It is likely a better use of a computational budget to restart a stagnant algorithm.

- Nothing in these tests prevent their inclusion in
  - Derivative-based algorithms
  - The refinement stage of global algorithms

- For further information, see:

  `http://www.mcs.anl.gov/~wild/tnoise`